# Data assimilation concepts and me March 1999

## By F. Bouttier and P. Courtier

Training Course Notes Front Page >>
Table of contents >>
Next Section >>
Previous Section >>



## 4 . Statistical interpolation with least-squares e

In this section we present the fundamental equation for linear analy form: the *least squares estimation*, also called *Best Linear Unbiase* following sections will provide more explanations and illustrations, least-squares estimation can be simplified to yield the most commo nowadays in meteorology and oceanography.

### 4.1 Notation and hypotheses

The dimension of the model state is $n$ and the dimension of the ob will denote:

$\mathbf{x}_t$ true model state (dimension $n$ )

$\mathbf{x}_b$ background model state (dimension $n$ )

$\mathbf{x}_a$ analysis model state (dimension $n$ )

$\mathbf{y}$ vector of observations (dimension $p$ )

$H$ observation operator (from dimension $n$ to $p$ )

$\mathbf{B}$ covariance matrix of the background errors $(\mathbf{x}_b - \mathbf{x}_t)$ (dime

$\mathbf{R}$ covariance matrix of observation errors $(\mathbf{y} - H[\mathbf{x}_t])$ (dimen

$\mathbf{A}$ covariance matrix of the analysis errors $(\mathbf{x}_a - \mathbf{x}_t)$ (dimensio

The following hypotheses are assumed:

- **Linearized observation operator**: the variations of the ob vicinity of the background state are linear: for any $\mathbf{x}$ close eno $H(\mathbf{x}) - H(\mathbf{x}_b) = \mathbf{H}(\mathbf{x} - \mathbf{x}_b)$ where $\mathbf{H}$ is a linear operator.
- **Non-trivial errors**: $\mathbf{B}$ and $\mathbf{R}$ are positive definite matrices.
- **Unbiased errors**: the expectation of the background and o i.e. $\overline{\mathbf{x}_b - \mathbf{x}_t} = \overline{\mathbf{y} - H(\mathbf{x}_t)} = 0$
- **Uncorrelated errors**: observation and background errors a i.e. $\overline{(\mathbf{x}_b - \mathbf{x}_t)(\mathbf{y} - H[\mathbf{x}_t])^T} = 0$
- **Linear analysis**: we look for an analysis defined by correc which depend linearly on background observation departures
- **Optimal analysis**: we look for an analysis state which is as true state in an r.m.s. sense (i.e. it is a minimum variance esti

ref: Daley 1991; Lorenc 1986; Ghil 1989

4.2 Theorem: least-squares analysis equations

(a) The *optimal least-squares estimator*, or *BLUE analysis*, following interpolation equations:

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - H[\mathbf{x}_b])$$

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}$$

where the linear operator $\mathbf{K}$ is called the *gain*, or *weight ma*

(a) The *analysis error covariance matrix* is, for any $\mathbf{K}$ :

$$\mathbf{A} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B}(\mathbf{I} - \mathbf{K}\mathbf{H})^T + \mathbf{K}\mathbf{R}\mathbf{K}^T$$

If $\mathbf{K}$ is the optimal least-squares gain, the expression becom

$$\mathbf{A} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B}$$

(a) The BLUE analysis is equivalently obtained as a solutio *optimization problem*:

$$\begin{aligned}\mathbf{x}_a &= \text{Arg min} J \\ J(\mathbf{x}) &= (\mathbf{x}-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + (\mathbf{y}-H[\mathbf{x}])^T\mathbf{R}^{-1}(\mathbf{y}- \\ &= J_b(\mathbf{x}) + J_o(\mathbf{x})\end{aligned}$$

where $J$ is called the *cost function* of the analysis (or *misfit* is the *background term*, $J_o$ is the *observation term*.

(a) The analysis $\mathbf{x}_a$ is *optimal*: it is closest in an r.m.s. sense

(b) If the background and observation error pdfs are Gauss *maximum likelihood estimator* of $\mathbf{x}_t$ . .

Proof:

With a translation of $\mathbf{x}$ by $\mathbf{x}_b$ , we can assume that $H = \mathbf{H}$ s operator is linear for our purposes. The equation (A1) is simp expression of the fact that we want the analysis to depend lin departures. The expression of $\mathbf{K}$ in (A2) is well-defined becau matrix, and $\mathbf{HBH}^T$ is positive. The minimization problem (A5) $J_o$ is a convex function and $J_b$ is a strictly convex function (it

The equivalence between items (a) and (c) of the theorem s requirement that the gradient of $J$ is zero at the optimum $\mathbf{x}_a$ :

$$\begin{aligned}\nabla J(\mathbf{x}_a) = 0 &= 2\mathbf{B}^{-1}(\mathbf{x}_a-\mathbf{x}_b) - 2\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{y}-H[\mathbf{x}_a]) \\ 0 &= \mathbf{B}^{-1}(\mathbf{x}_a-\mathbf{x}_b) - \mathbf{H}^T\mathbf{R}^{-1}(\mathbf{y}-H[\mathbf{x}_b]) - \mathbf{H}^T \\ (\mathbf{x}_a-\mathbf{x}_b) &= (\mathbf{B}^{-1}+\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{y}-H[\mathbf{x}_b])\end{aligned}$$

The identity with (A2) is straightforward to prove (all inverse positive definite):

$$\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{B}\mathbf{H}^T+\mathbf{R}) = (\mathbf{B}^{-1}+\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})\mathbf{B}\mathbf{H}^T$$
$$= \mathbf{H}^T+\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}\mathbf{H}^T$$

hence

$$(\mathbf{B}^{-1}+\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T+\mathbf{R})^{-}$$

The expressions (A3) and (A4) for $\mathbf{A}$ are obtained by rewriti (A1) in terms of the background, analysis and observation err

$$\varepsilon_b = \mathbf{x}_b - \mathbf{x}_t$$
$$\varepsilon_a = \mathbf{x}_a - \mathbf{x}_t$$
$$\varepsilon_o = \mathbf{y} - H[\mathbf{x}_t]$$
$$\varepsilon_a - \varepsilon_b = \mathbf{K}(\varepsilon_o - \mathbf{H}\varepsilon_b)$$
$$\varepsilon_a = (\mathbf{I}-\mathbf{K}\mathbf{H})\varepsilon_b + \mathbf{K}\varepsilon_o$$

By developing the expression of $\varepsilon_a\varepsilon_a^T$ and taking its expectat expectation operator one finds the general expression (A3) (r being uncorrelated, their cross-covariance is zero). The simpl derive by substituting the expression for the optimal $\mathbf{K}$ and si cancel.

Finally to prove (A2) itself we take the analysis error covaria and we minimize its trace, i.e. the total error variance: (note th

$$\mathrm{Tr}(\mathbf{A}) = \mathrm{Tr}(\mathbf{B}) + \mathrm{Tr}(\mathbf{K}\mathbf{H}\mathbf{B}\mathbf{H}^T\mathbf{K}^T) - 2\mathrm{Tr}(\mathbf{B}\mathbf{H}^T\mathbf{K}^T) + \mathrm{Tr}$$

This is a continuous differentiable scalar function of the coef express its derivative $d_{\mathbf{K}}$ as the first- order terms in $\mathbf{K}$ of the d $\mathrm{Tr}(\mathbf{A})(\mathbf{K}+\mathbf{L}) - \mathrm{Tr}(\mathbf{A})(\mathbf{K})$ , $\mathbf{L}$ being an arbitrary test matrix:

$$d_{\mathbf{K}}[\mathrm{Tr}(\mathbf{A})]\mathbf{L} = 2\mathrm{Tr}(\mathbf{KHBH}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}) - 2\mathrm{Tr}(\mathbf{BH}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}) + 2\mathrm{T}:$$
$$= 2\mathrm{Tr}(\mathbf{KHBH}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}} - \mathbf{BH}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}} + \mathbf{KRL}^{\mathrm{T}})$$
$$= 2\mathrm{Tr}\{[\mathbf{K}(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R}) - \mathbf{BH}^{\mathrm{T}}]\mathbf{L}^{\mathrm{T}}\}$$

The last line shows that the derivative is zero for any choice $(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})\mathbf{K}^{\mathrm{T}} - \mathbf{HB} = 0$ , which is equivalent to

$$\mathbf{K} = \mathbf{BH}^{\mathrm{T}}(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})^{-1}$$

because $(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})$ is assumed to be invertible.

In the case of Gaussian pdfs, one can model the background, obse as follows, respectively: ($b$ , $o$ and $a$ are normalization factors.)

$$\mathcal{P}_{\mathrm{b}}(\mathbf{x}) = b \exp\left[\frac{1}{2}(\mathbf{x} - \mathbf{x}_{\mathrm{b}})^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_{\mathrm{b}})\right]$$

$$\mathcal{P}_{\mathrm{o}}(\mathbf{x}) = o \exp\left[\frac{1}{2}(\mathbf{y} - H[\mathbf{x}])^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y} - H[\mathbf{x}_{\mathrm{b}}])\right]$$

$$\mathcal{P}_{\mathrm{a}}(\mathbf{x}) = \mathcal{P}_{\mathrm{b}}(\mathbf{x})\mathcal{P}_{\mathrm{o}}(\mathbf{x})$$

which yields the right averages and covariances for the backgroun and the analysis error pdf is simply defined as the Bayesian produc sources of information, the background and the observation pdfs (t rigorously by using Bayes' theorem to write $\mathcal{P}_{\mathrm{a}}$ as a conditional pro observations and the a priori pdf of the background). Then, by takin $\mathcal{P}_{\mathrm{a}}(\mathbf{x})$ , one finds that the model state with the maximum probability that minimizes the cost function $J(\mathbf{x})$ expressed in the theorem.

### 4.3 Comments

The hypotheses of non-triviality can always been made in well-pos is non-positive, one can restrict the control space to the orthogonal analysis will not make any correction to background variables that is not a surjection, then some observations are redundant and the be restricted to the image of $\mathbf{H}$ . If $\mathbf{R}$ is non-positive, the expressio (then the analysis will be equal to the observed value at the observ

the variational version of the least-squares analysis cannot be use some algebraic precautions) to have some infinite eigenvalues in J which means that some observations are not used because their e

The hypothesis of *unbiased errors* is a difficult one in practice beca significant biases in the background fields (caused by biases in the the observations (or in the observation operators). If the biases are subtracted from the background and observation values, and the a the debiased quantities. If the biases are left in, the analysis will no it will seem to reduce the biases by interpolating between the back It is important to monitor the biases in an assimilation system, e.g. background departures, but it is not trivial to decide which part of th observation biases. The problem of bias monitoring and removal is research.

The hypothesis of *uncorrelated errors* is usually justified because t background and in the observations are supposed to be completel one must be careful about observation preprocessing practices (su procedures) that use the background field in a way that biases the background. It might reduce the apparent background departures, analysis to be suboptimal (too close to the background, a condition *problem*).

The *tangent linear hypothesis* is not trivial and it is commented in t

It is possible to rewrite the least-squares analysis equations in term error covariance matrices, called *information* matrices. It makes the complicated, but it allows one to see clearly that the *information* co the sum, in a simple sense, of the observations provided by the ba observations. This is illustrated in the section on the estimation of a

It will be shown in the section on dual algorithms (PSAS analysis) t particular the cost function $J$ , can be rewritten in the space of the easy to that least-squares analysis is closely related to a linear reg state and observations.

### 4.4 On the tangent linear hypothesis

The hypothesis of *linearized observation operator* is needed in ord algebraic expression for the optimal $\mathbf{K}$ . In practice, $H$ may not be makes physical sense to linearize it in the vicinity of the backgroun

$$H(\mathbf{x}) - H(\mathbf{x}_b) \approx \mathbf{H}(\mathbf{x} - \mathbf{x}_b)$$

Then, $\mathbf{K}$ being a continuous function of $\mathbf{H}$ , the least-squares equa should intuitively yield a nearly optimal $\mathbf{x}_a$ .

More generally, the *tangent linear hypothesis* on $H$ can be written Young formula in the vicinity of an arbitrary state $\mathbf{x}$ and for a pertur

$$H(\mathbf{x}+h) = H(\mathbf{x}) + \mathbf{H}h + \mathrm{O}(\|h\|^2)$$

with $\lim_{h \to 0} \mathrm{O}(\|h\|^2)h^{-2} = 0$ . This hypothesis, called the *tangent lin* acceptable if the higher-order variations of $H$ can be neglected (in no discontinuities) for all perturbations of the model state which ha magnitude as the background errors. The operator $\mathbf{H}$ is called the *derivative*, or *tangent linear (TL)*[1] function of $H$ at point $\mathbf{x}$ . Althoug mathematical property of $H$ , it is not enough for practical purposes approximation

$$H(\mathbf{x}+h) - H(\mathbf{x}) \approx \mathbf{H}h$$

must be satisfactory, in user-defined terms, for finite values of $h$ th application considered. In the least-squares analysis problem, we n

$$\mathbf{y} - H(\mathbf{x}) \approx \mathbf{y} - \mathbf{H}(\mathbf{x} - \mathbf{x}_b) + H(\mathbf{x}_b)$$

for all values of $\mathbf{x}$ that will be encountered in the analysis procedur and also all trial values used in the minimization of $J(\mathbf{x})$ if a variatio performed[2]. Thus the important requirement is that the difference b [x] should be much smaller than the typical observation erro model state perturbations $\mathbf{x} - \mathbf{x}_b$ of size and structure consistent wi errors, and also with the amplitude of the analysis increments $\mathbf{x}_a -$

Thus the problem of linearizing $H$ is not just related to the observa must be appreciated in terms of the errors in the background $\mathbf{x}_b$ too assimilation system are the previous forecast errors, which depend and the quality of the model. Ultimately the correctness of the linea appreciated in the context of the fully integrated assimilation system the linearization to a good system because the departures $\mathbf{x} - \mathbf{x}_b$ w the linearization may be inapplicable to difficult data assimilation pr case with ocean models or satellite data, which means that it can b sophisticated analysis algorithms that rely too much on the linearity

The linearization problem can be even more acute for the lineariza operator $M$ which is needed in 4D-Var and in the Kalman filter des linearization of $H$ , it may or may not be licit depending on the qua the assimilation system: data coverage, observation quality, model and forecast range. The user requirements and the physical prope be considered.

**The non-linear analysis problem**

The assumption of linear analysis is a strong one. Linear algebra is optimal analysis equations. One can rely on the linearization of a w observation operator, at the expense of optimality. The *incrementa* for the variational analysis) performs this procedure iteratively in an make the analysis more optimal. For strongly non-linear problems, simple way to calculate the optimal analysis. The *simulated annea* specific methods, such as the *simplex*, deal with variables with bou Finally, it is sometimes possible to make a problem more linear sim of model and observation variables (see the section on minimizatio

**4.5 The point of view of conditional probabilities**

It is interesting to formalize the analysis problem using the conditio probabilities. Let us denote $P(x)$ the a priori pdf (probability density state before the observations are considered, i.e. the background p the pdf of the observations. The aim of the analysis is to find the m conditional probability of the model state given the observations. T (i.e. the probability that $x$ and $y$ occur together) is

$$P(x \wedge y) = P(x|y)P(y) = P(y|x)P(x)$$

i.e. it is the probability that $x$ occurs when $y$ occurs, and vice versa the Bayes theorem. In the analysis procedure we know that a mea and we know its value $y$ , so $P(y) = 1$ and we obtain

$$P(x|y) = P(y|x)P(x)$$

which means that the analysis pdf is equal to the background pdf t $P(y|x)$ . The latter peaks at $y = H(x)$ but it is not a Dirac distributi observations are not error-free.

The virtue of the probabilistic derivation of the analysis problem is t non-Gaussian probabilities (although this spoils the equivalence w $\mathbf{K}$ ). A practical application is done in the framework of variational q

assumed that observation errors are not Gaussian but they contain errors", i.e. there is a probability that the error is not generated by t physical processes but by some more serious problem, like coding The gross errors might be modelled using a uniform pdf over a pre admissible gross errors, leading to a non-Gaussian observation pd the logarithm of this pdf is taken, the resulting observation cost fun gives less weight to the observation (i.e. there is less slope) for mo strongly with the observed value.

ref: Lorenc 1986

### 4.6 Numerical cost of least-squares analysis

In current operational meteorological models, the dimension of the precisely, of the control variable space) $\mathbf{x}$ is of the order of $n = 10$ the observation vector (the number of observed scalars) is of the o analysis[3]. Therefore the analysis problem is mathematically underd some regions it might be overdetermined if the density of the obser resolution of the model). In any practical application it is essential t the matrix operators involved in computing the analysis (Fig. 4 ). Th method requires in principle the specification of covariance matrice inverses in the variational form of the algorithm) which respectively $n^2/2$ and $p^2/2$ distinct coefficients, which are statistics to estimate variance or covariance statistic converges like the square root of th The explicit determination of $\mathbf{K}$ requires the inversion of a matrix o asymptotic complexity of the order of $p^2 \log(n)$ . The exact minimiz $J$ requires, in principle, $n+1$ evaluations of the cost function and i quadratic and there are no numerical errors (e.g. using a conjugate

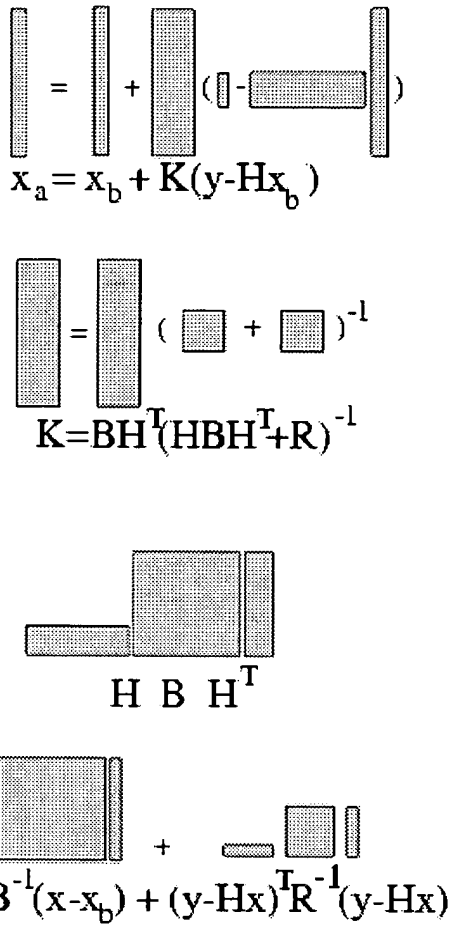


***Figure 4 . Sketches of the shapes of the matrices and vector d an usual analysis problem where there are many fewer observ freedom in the model: from top to bottom, in the equations of computation of*** $\mathbf{K}$ ***, of the*** $\mathbf{HBH}^T$ ***term, and the computation of***

It is obvious that, except in analysis problems of very small dimens retrievals), it is impossible to compute exactly the least-squares an approximations are necessary, they are the subject of the following

### 4.7 Conclusion

We have seen that there are two main ways of defining the statistic

- either assume that the background and error covariances a analysis equations by requiring that the total analysis error va
- or assume that the background and observation error pdfs the analysis equations by looking for the state with the maxim

Both approaches lead to two mathematically equivalent algorithms

- the direct determination of the analysis gain matrix $\mathbf{K}$ ,

- the minimization of a quadratic cost function.

These algorithms have very different numerical properties, and the soon as some underlying hypotheses are not verified, like the linea operator, for instance.

---

**Training Course Notes Front Page >>**
**Table of contents >>**
**Next Section >>**
**Previous Section >>**

---

[1] Both qualifiers *tangent* and *linear* are needed: obviously $\mathbf{H}$ could satisfying the Taylor formula. A function can also be tangent to ano if the difference between them is an $o(\|h\|^2)$ , e.g. $x^2$ and $x^3$ are tan $x = 0$ .

[2] Qualitatively speaking they all belong to a neighbourhood of $\mathbf{x}_b$ h which is consistent with the $\mathbf{B}$ and $\mathbf{R}$ error covariances.

[3] At ECMWF in winter 1998 the control variable dimension was 51 observations (per 6-hour interval) was about 150000

03.12.2001